

SUPERLINEAR CONVERGENCE OF AN INTERIOR-POINT METHOD DESPITE DEPENDENT CONSTRAINTS

DANIEL RALPH* AND STEPHEN J. WRIGHT†

Abstract. We show that an interior-point method for monotone variational inequalities exhibits superlinear convergence provided that all the standard assumptions hold except for the well-known assumption that the Jacobian of the active constraints has full rank at the solution. We show that superlinear convergence occurs even when the constant-rank condition on the Jacobian assumed in an earlier work does not hold.

AMS(MOS) subject classifications. 90C33, 90C30, 49M45

1. Introduction. We consider the following monotone variational inequality over a closed convex set $\mathcal{C} \subset \mathbb{R}^n$:

$$(1) \quad \text{Find } z \in \mathcal{C} \text{ such that } (z' - z)^T \Phi(z) \geq 0, \quad \text{for all } z' \in \mathcal{C},$$

where $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and the set \mathcal{C} is defined by the following algebraic inequality:

$$\mathcal{C} = \{z \mid g(z) \leq 0\},$$

where $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$. The mapping Φ is assumed to be C^1 (continuously differentiable) and monotone; that is,

$$(z' - z)^T (\Phi(z') - \Phi(z)) \geq 0 \quad \text{for all } z', z \in \mathbb{R}^n,$$

while each component function $g_i(\cdot)$ of $g(\cdot)$ is convex and twice continuously differentiable.

By introducing $g(\cdot)$ explicitly into the problem (1), we obtain the following mixed nonlinear complementarity (NCP) problem: Find the vector triple $(z, \lambda, y) \in \mathbb{R}^{n+2m}$ such that

$$(2) \quad \begin{bmatrix} 0 \\ y \end{bmatrix} = \begin{bmatrix} f(z, \lambda) \\ -g(z) \end{bmatrix}, \quad (\lambda, y) \geq 0, \quad \lambda^T y = 0,$$

where $f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^n$ is the C^1 function defined by

$$(3) \quad f(z, \lambda) = \Phi(z) + Dg(z)^T \lambda.$$

It is well known [5] that, under suitable conditions on g such as the Slater constraint qualification, z solves (1) if and only if there exists a multiplier λ such that (z, λ) solves (2).

To show superlinear (local) convergence in methods for nonlinear programs, one usually makes several assumptions with regard to the solution point. Until recently, these assumptions included (local) uniqueness of the solution (z, λ, y) . This uniqueness condition was relaxed somewhat in [10] to allow for several multipliers λ corresponding

* Department of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria 3052, Australia. The work of this author was supported by the Australian Research Council.

† Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, U.S.A. This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Computational and Technology Research, U.S. Department of Energy, under Contract W-31-109-Eng-38.

to a locally unique solution z^* of (1), by introducing a constant-rank condition on the gradients of the constraints g_i that are active at z^* . The point of this article is to show that superlinear convergence holds in the previous setting [10] even when the constant-rank condition does not hold. This result lends theoretical support to our numerical observations [10, Section 7].

Briefly stated, the assumptions we make to obtain the superlinear results are as follows: monotonicity and differentiability of the mapping from (z, λ) to $(f(z, \lambda), -g(z))$, such that the partial derivative with respect to z is Lipschitz near z^* ; a positive definiteness condition to ensure invertibility of the linear system that is solved at each iteration of the interior-point method; the Slater constraint qualification on g ; existence of a strictly complementary solution; and a second-order condition that guarantees local uniqueness of the solution z^* of (1). A formal statement of these assumptions and further details are given in Section 2.2.

Before the present paper was written, superlinear convergence had been proved for other methods for nonlinear programming without the strict complementarity assumption, but these results typically required the Jacobian of active constraints to have full rank (see Pang [8], Bonnans [1], and Facchinei, Fischer, and Kanzow [2]). Monteiro and Zhou [7] undertook an alternative line of investigation, namely, primal-dual interior-point methods for linearly constrained convex programs where the objective function satisfied a relaxed second-order condition that allowed for multiple (primal) solutions. They show superlinear convergence of the method of Ralph and Wright [10] for this class of problems.

Since the first version this paper was written, there have been several algorithmic developments for which superlinear convergence can be proved under conditions that do not imply linear independence of the active constraint gradients, or indeed multiplier nondegeneracy, at the solution point. Fischer [3] modifies the classical sequential quadratic programming (SQP) method of Wilson to produce a method that is quadratically convergent under some nonstandard assumptions that do not imply multiplier uniqueness. Qi and Wei [9] show superlinear convergence of a feasible point SQP algorithm in the presence of multiplier degeneracy by requiring, amongst other things, a kind of constant-rank condition on the constraint gradients that are active at the optimal solution. Wright discusses SQP in [13, 14] and interior-point methods in [15]. The paper [13] presents a stabilized version of SQP that exhibits quadratic convergence while allowing for multiplier degeneracy under strict complementarity and otherwise standard conditions; see Hager [4] for a convergence analysis of this method without strict complementarity. The recent paper [15] uses linear algebra arguments to show, among other things, that local superlinear convergence can be obtained for nonlinear programs under assumptions similar to those discussed here, except that one can omit the monotonicity (convexity) condition (Assumption 1 in Section 2.2).

Possibly the best known application of (1) is the convex programming problem defined by

$$(4) \quad \min_z \phi(z) \quad \text{subject to} \quad z \in \mathcal{C},$$

where $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is C^2 and convex. Let $\Phi = D\phi$. It is easy to show that the NCP formulation (2),(3) is equivalent to the standard Karush-Kuhn-Tucker (KKT) conditions for (4). If a constraint qualification holds, then solutions of (4) correspond, via Lagrange multipliers, to solutions of (2)–(3) and, in addition, solutions of (1) and (4) coincide.

We consider the solution of (1) by the interior-point algorithm of Ralph and

Wright [10], which is in turn a natural extension of the safe-step/fast-step algorithm of Wright [11] for monotone linear complementarity problems. The algorithm is based on a restatement of the problem (2) as a set of constrained nonlinear equations, as follows:

$$(5) \quad \begin{bmatrix} -f(z, \lambda) \\ y + g(z) \\ -\Lambda Y e \end{bmatrix} = \begin{bmatrix} r_f(z, \lambda) \\ r_g(z, y) \\ -\Lambda Y e \end{bmatrix} = 0, \quad (\lambda, y) \geq 0,$$

where the residuals r_f and r_g are defined in an obvious way. All iterates (z^k, λ^k, y^k) satisfy the positivity conditions *strictly*; that is, $(\lambda^k, y^k) > 0$ for all $k = 0, 1, 2, \dots$. The interior-point algorithm can be viewed as a modified Newton's method applied to the equality conditions in (5), in which search directions and step lengths are chosen to maintain the positivity condition on (λ, y) . Near a solution, the algorithm takes steps along the pure Newton direction defined by

$$(6) \quad \begin{bmatrix} D_z f & Dg^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta \lambda \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_f(z, \lambda) \\ r_g(z, y) \\ -\Lambda Y e \end{bmatrix}.$$

The solution $(\Delta z, \Delta \lambda, \Delta y)$ of this system is also known as the affine-scaling direction.

The duality measure defined by

$$\mu = \lambda^T y / m$$

is used frequently in our analysis as a measure of noncomplementarity and infeasibility.

To extend the superlinear convergence result of [10] without a constant-rank condition on the active constraint Jacobian, we show that the affine-scaling step defined by (6) has size $O(\mu)$. Hence, the superlinearity result can be extended to most algorithms that take near-unit steps along directions that are asymptotically the same as the affine-scaling direction.

Since we are extending our work in [10], much of the analysis in the earlier work carries over without modification to the present case, and we omit many of the details here. We focus instead on the main technical result needed to prove fast local convergence—the estimate $(\Delta z, \Delta \lambda, \Delta y) = O(\mu)$ for the affine-scaling step—and restate just enough of the earlier material to make the current note self-contained.

2. The Algorithm and its Convergence. In this section, we review the notation and assumptions of the algorithm from Ralph and Wright [10]. We also state the main global and superlinear convergence results, which differ from the corresponding theorems in [10] only in the absence of the constant-rank assumption. Using the framework of [10], we can accomplish our main goal—a proof of Theorem 2.2—without referring directly to the algorithm. The statement of the algorithm is given in the appendix for completeness.

2.1. Notation and Terminology. We use \mathcal{S} to denote the solution set for (2), and $\mathcal{S}_{z, \lambda}$ to denote its projection onto its first $n + m$ components; that is,

$$\mathcal{S} = \{(z, \lambda, y) \mid (z, \lambda, y) \text{ solves (2)}\}, \quad \mathcal{S}_{z, \lambda} = \{(z, \lambda) \mid (z, \lambda, -g(z)) \in \mathcal{S}\}.$$

For a particular z^* to be defined in Assumption 4 in the next subsection, we define

$$(7) \quad \mathcal{S}_\lambda^* = \{\lambda \mid (z^*, \lambda) \in \mathcal{S}_{z, \lambda}\}.$$

We can partition $\{1, 2, \dots, m\}$ into basic and nonbasic index sets \mathcal{B} and \mathcal{N} such that for all solutions $(z^*, \lambda^*, y^*) \in \mathcal{S}$, we have

$$\lambda_i^* = 0, \text{ for all } i \in \mathcal{N}; \quad y_i^* = 0, \text{ for all } i \in \mathcal{B}.$$

The solution $(z^*, \lambda^*, -g(z^*))$ is *strictly complementary* if $\lambda^* + y^* > 0$; that is, $\lambda_i^* > 0$ for all $i \in \mathcal{B}$ and $y_i^* = -g(z^*) > 0$ for all $i \in \mathcal{N}$.

We use $\lambda_{\mathcal{N}}$ and $\lambda_{\mathcal{B}}$ to denote the subvectors of λ that correspond to the index sets \mathcal{N} and \mathcal{B} , respectively. Similarly, we use $Dg_{\mathcal{B}}(z)$ to denote the $|\mathcal{B}| \times n$ row submatrix of $Dg(z)$ corresponding to \mathcal{B} .

Finally, if we do not specify the arguments for functions g , Dg , f , and so on, they are understood to be the appropriate components of the current point (z, λ, y) . The notation Dg^* refers to $Dg(z^*)$.

2.2. Assumptions. Here we give a formal statement of the assumptions needed for global and superlinear convergence. Some motivation is given here, but we refer the reader to the earlier paper [10] for further details.

The first assumption ensures that the mapping f defined by (3) is monotone with respect to z and therefore that the mapping $(z, \lambda) \rightarrow (f(z, \lambda), -g(z))$ is monotone.

ASSUMPTION 1. $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is C^1 and monotone; and each component function g_i of $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is C^2 and convex.

The second assumption requires positive definiteness of a certain matrix projection, to ensure that the coefficient matrix of the Newton-like system to be solved for each step in the interior-point algorithm is nonsingular (see the appendix, (48)).

ASSUMPTION 2. *The two-sided projection of the matrix*

$$D_z f(z, \lambda) = D\Phi(z) + \sum_{i=1}^m \lambda_i D^2 g_i(z)$$

onto $\ker Dg(z)$ is positive definite for all $z \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}_{++}^m$; that is, for any basis matrix Z of $\ker Dg(z)$, the matrix $Z^T D_z f(z, \lambda) Z$ is positive definite.

Note that this assumption is trivially satisfied when the nonnegativity condition $z \geq 0$ is incorporated in the constraint function $g(\cdot)$.

We assume, too, that the Slater condition holds for the constraint function g .

ASSUMPTION 3. *There is a vector $\bar{z} \in \mathcal{C}$ such that $g(\bar{z}) < 0$.*

Next, we assume the existence (but not uniqueness) of a strictly complementary solution.

ASSUMPTION 4. *There is a strictly complementary solution (z^*, λ^*, y^*) , that is, (z^*, λ^*, y^*) satisfies (2) with $\lambda^* + y^* > 0$.*

The strict complementarity condition is essential for superlinear convergence in a number of contexts besides NCP and nonlinear programming. See, for example Wright [12, Chapter 7] for an analysis of linear programming and Monteiro and Wright [6] for asymptotic properties of interior-point methods for monotone linear complementarity problems.

Next, we make a smoothness assumption on Φ and g in the neighborhood of the first component z^* of the strictly complementary solution from Assumption 4.

ASSUMPTION 5. *The matrix-valued functions $D\Phi$ and $D^2 g_i$, $i = 1, 2, \dots, m$ are Lipschitz continuous in a neighborhood of z^* .*

Finally, we make an invertibility assumption on the projection of the Hessian onto the kernel of the active constraint Jacobian. This assumption is essentially a second-order sufficient condition for optimality.

ASSUMPTION 6. *Let z^* be defined as in Assumption 4, and let \mathcal{B} , $\mathcal{S}_{z,\lambda}$ and \mathcal{S}_λ^* be defined as in Section 2. Then for each $\lambda \in \mathcal{S}_\lambda^*$, the two-sided projection of $D_z f(z^*, \lambda)$ onto $\ker(Dg_{\mathcal{B}}^*)$ is invertible.*

We mention that with Assumption 2, the invertibility condition of Assumption 6 is equivalent to positive definiteness of the two-sided projection.

We show in [10, Lemma 4.2] that, under the assumptions above, the first component of all solutions $(z, \lambda) \in \mathcal{S}_{z,\lambda}$ is z^* , so that $\mathcal{S}_{z,\lambda}$ has the form $\mathcal{S}_{z,\lambda} = \{z^*\} \times \mathcal{S}_\lambda^*$. Moreover, Lemma 4.1 of [10] shows that the set \mathcal{S}_λ^* defined in (7) is polyhedral, convex, and compact, and is therefore equal to the convex hull of its extreme points.

In the statements of our results, we refer to a set of ‘‘standing assumptions,’’ which we define as follows:

Standing Assumptions: Assumptions 1–6, together with an assumption that the algorithm of Ralph and Wright [10] applied to the problem (2) generates an infinite sequence $\{(z^k, \lambda^k, y^k)\}$ with a limit point.

Along with Assumptions 1–6, the superlinear convergence result in Ralph and Wright [10] requires a constant-rank constraint qualification to hold. To be specific, the analysis of that paper requires the existence of an open neighborhood U of z^* such that for all matrix sequences $\{H^k\} \subset \{Dg_{\mathcal{B}}(z)^T \mid z \in U\}$ with $H^k \rightarrow H^* = Dg_{\mathcal{B}}(z^*)^T$ and all index sets $\mathcal{J} \subset \{1, 2, \dots, |\mathcal{B}|\}$, we have that

$$\text{rank } H^k_{\mathcal{J}} \rightarrow \text{rank } H^*_{\mathcal{J}}.$$

However, in the analysis of [10], this assumption is not invoked until Section 5.4, so we are justified in reusing many results from earlier sections of that paper here. Indeed, we also reuse results from later sections of [10] by applying them to *constant* matrices (which certainly satisfy the constant-rank condition).

The algorithm makes use of a family of sets $\Omega(\gamma, \beta)$ defined for positive parameters γ and β as follows:

$$(8) \quad \begin{aligned} \Omega(\gamma, \beta) = \{ & (z, \lambda, y) \mid (\lambda, y) \geq 0, \|r_f(z, \lambda)\| \leq \beta\mu, \\ & \|r_g(z, y)\| \leq \beta\mu, \lambda_i y_i \geq \gamma\mu, \quad i = 1, 2, \dots, m\}. \end{aligned}$$

In particular, the k th iterate (z^k, λ^k, y^k) belongs to $\Omega(\gamma_k, \beta_k)$, where the algorithm chooses the sequences $\{\gamma_k\}$ and $\{\beta_k\}$ to satisfy

$$\begin{aligned} 0 < \beta_{\min} = \beta_0 & \leq \beta_1 \leq \dots \leq \beta_k \leq \dots < \beta_{\max}, \\ \gamma_{\max} = \gamma_0 & \geq \gamma_1 \geq \dots \geq \gamma_k \geq \dots \geq \gamma_{\min} > 0. \end{aligned}$$

Given the notation

$$\Omega_k \triangleq \Omega(\gamma_k, \beta_k), \quad \Omega \triangleq \Omega(\gamma_{\min}, \beta_{\max}),$$

it is easy to see that

$$\Omega_0 \subset \Omega_1 \subset \dots \subset \Omega_k \subset \dots \subset \Omega.$$

Since all iterates (z^k, λ^k, y^k) belong to Ω , and since the residual norms $\|r_f\|$ and $\|r_g\|$ are bounded in terms of μ for vectors in this set, we are justified in using μ alone as an indicator of progress, rather than a merit function that also takes account of the residual norms.

We have already noted that z^* is the first component of all solutions, under the standing assumptions, and that \mathcal{S}_λ^* is compact. Moreover, we show in [10, Theorem 3.2] (see also Theorem 2.1 below) that if the our algorithm generates an infinite sequence (z^k, λ^k, y^k) , all its limit points lie in \mathcal{S} . Let $(z^*, \hat{\lambda}, y^*)$ be any fixed limit point; it follows immediately that

$$(9) \quad (z^*, \hat{\lambda}, y^*) \in \mathcal{S},$$

where $y^* = -g(z^*)$ by definition. We are particularly interested in points in Ω that lie close to this limit point, so we consider the near-solution neighborhood $\mathcal{S}(\delta)$ defined by

$$(10) \quad \mathcal{S}(\delta) \triangleq \{(z, \lambda, y) \in \Omega \mid \|(z, \lambda, y) - (z^*, \hat{\lambda}, y^*)\| \leq \delta\}.$$

2.3. Convergence of the Algorithm. The algorithm converges globally according to the following theorem.

THEOREM 2.1. (Ralph and Wright [10, Theorem 3.2]) *Suppose that Assumptions 1 and 2 hold. Then either*

- (A) $(z^k, \lambda^k, y^k) \in \mathcal{S}$ for some $k < \infty$, or
- (B) *all limit points of $\{(z^k, \lambda^k, y^k)\}$ belong to \mathcal{S} .*

Here, however, our focus is on local superlinear convergence. The theorem below is simply a restatement of [10, Theorem 3.3] *without* the constant-rank condition on the active constraint Jacobian matrix [10, Assumption 7].

THEOREM 2.2. *Suppose that Assumptions 1, 2, 3, 4, 5, and 6 are satisfied and that the sequence $\{(z^k, \lambda^k, y^k)\}$ is infinite, with a limit point $(z^*, \hat{\lambda}, y^*) \in \mathcal{S}$. Then the algorithm eventually always takes fast steps, and*

- (i) *the sequence $\{\mu_k\}$ converges superlinearly to zero with Q -order at least $1 + \hat{\tau}$, and*
- (ii) *the sequence $\{(z^k, \lambda^k, y^k)\}$ converges superlinearly to $(z^*, \hat{\lambda}, y^*)$ with R -order at least $1 + \hat{\tau}$.*

The proof of this result follows that of the earlier paper in all respects except for the estimate

$$(11) \quad (\Delta z, \Delta \lambda, \Delta y) = O(\mu)$$

for the affine-scaling step calculated from (6). The remainder of the paper is devoted to proving that this estimate holds under the given assumptions.

3. An $O(\mu)$ Estimate for the Affine-Scaling Step. Our strategy for proving the estimate (11) for the step (6) is based on a partitioning of the right-hand side in (6). The following vectors are useful in defining the partition.

$$\begin{aligned} (12a) \quad \eta_f &= D_z f(z, \lambda)(z^* - z) + Dg(z)^T(\bar{\pi} - \lambda), \\ (12b) \quad \eta_g &= y - Dg(z)(z^* - z) + g(z^*), \\ (12c) \quad \bar{\eta}_f &= D_z f(z, \lambda)(z^* - z) + Dg(z^*)^T(\bar{\pi} - \lambda), \\ (12d) \quad \bar{\eta}_g &= y - Dg(z^*)(z^* - z) + g(z^*), \\ (12e) \quad \epsilon_f &= -f(z, \lambda) - D_z f(z, \lambda)(z^* - z) - Dg(z)^T(\bar{\pi} - \lambda), \\ (12f) \quad \epsilon_g &= g(z) - g(z^*) + Dg(z)(z^* - z), \end{aligned}$$

where z^* is defined in Assumption 4 and $(z^*, \bar{\pi})$ is the projection of the current point (z, λ) onto the set $\mathcal{S}_{z, \lambda}$ of (z, λ) solution components. The right-hand side of (6) can

be partitioned as

$$\begin{bmatrix} r_f \\ r_g \\ -\Lambda Y e \end{bmatrix} = \begin{bmatrix} \eta_f \\ \eta_g \\ -\Lambda Y e \end{bmatrix} + \begin{bmatrix} \epsilon_f \\ \epsilon_g \\ 0 \end{bmatrix}.$$

We define a corresponding splitting of the affine-scaling step:

$$(13) \quad (\Delta z, \Delta \lambda, \Delta y) = (t, u, v) + (t', u', v'),$$

where (t, u, v) and (t', u', v') satisfy the following linear systems:

$$(14) \quad \begin{bmatrix} D_z f & (Dg)^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} t \\ u \\ v \end{bmatrix} = \begin{bmatrix} \eta_f \\ \eta_g \\ -\Lambda Y e \end{bmatrix},$$

$$(15) \quad \begin{bmatrix} D_z f & (Dg)^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} t' \\ u' \\ v' \end{bmatrix} = \begin{bmatrix} \epsilon_f \\ \epsilon_g \\ 0 \end{bmatrix}.$$

We define a third variant on (6) as follows:

$$(16) \quad \begin{bmatrix} D_z f & (Dg^*)^T & 0 \\ -(Dg^*) & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \widehat{\Delta z} \\ \widehat{\Delta \lambda} \\ \widehat{\Delta y} \end{bmatrix} = \begin{bmatrix} \bar{\eta}_f \\ \bar{\eta}_g \\ -\Lambda Y e \end{bmatrix},$$

and split the step $(\widehat{\Delta z}, \widehat{\Delta \lambda}, \widehat{\Delta y})$ as

$$(17) \quad (\widehat{\Delta z}, \widehat{\Delta \lambda}, \widehat{\Delta y}) = (\tilde{t}, \tilde{u}, \tilde{v}) + (\tilde{t}', \tilde{u}', \tilde{v}'),$$

where $(\tilde{t}, \tilde{u}, \tilde{v})$ and $(\tilde{t}', \tilde{u}', \tilde{v}')$ satisfy

$$(18) \quad \begin{bmatrix} D_z f & (Dg^*)^T & 0 \\ -(Dg^*) & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \tilde{t} \\ \tilde{u} \\ \tilde{v} \end{bmatrix} = \begin{bmatrix} \bar{\eta}_f \\ \bar{\eta}_g \\ 0 \end{bmatrix}$$

$$(19) \quad \begin{bmatrix} D_z f & (Dg^*)^T & 0 \\ -(Dg^*) & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \tilde{t}' \\ \tilde{u}' \\ \tilde{v}' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\Lambda Y e \end{bmatrix}.$$

Because of Assumption 2, the matrices in (14), (15), (16), (18), and (19) are all invertible, so all these systems have unique solutions.

Our basic strategy for proving the estimate (11) is as follows. From [10, Section 5.3], we have without assuming the constant-rank condition that $(t', u', v') = O(\mu)$ for all $(z, \lambda, y) \in \mathcal{S}(\delta)$, where $\delta \in (0, 1)$ is a positive constant. The constant-rank assumption is, however, needed in [10] to prove that the other step component (t, u, v) is also $O(\mu)$. In this article, we obtain the same estimate *without* the constant-rank assumption, by proving that

$$(20) \quad (\widehat{\Delta z}, \widehat{\Delta \lambda}, \widehat{\Delta y}) = O(\mu), \quad (\widehat{\Delta z} - t, \widehat{\Delta \lambda} - u, \widehat{\Delta y} - v) = O(\mu),$$

for all $(z, \lambda, y) \in \mathcal{S}(\delta)$.

Our first result, proved in the earlier paper [10], collects some bounds that are useful throughout this section.

LEMMA 3.1. [10, Lemma 5.1] *Suppose that the standing assumptions hold. Then there is a constant C_0 such that the following bounds hold for all $(z, \lambda, y) \in \mathcal{S}(1)$:*

$$\begin{aligned} (21a) \quad & \lambda_i \leq C_0 \mu \quad (i \in \mathcal{N}), & y_i & \leq C_0 \mu \quad (i \in \mathcal{B}), \\ (21b) \quad & \lambda_i \geq \gamma_{\min}/C_0 \quad (i \in \mathcal{B}), & y_i & \geq \gamma_{\min}/C_0 \quad (i \in \mathcal{N}), \\ (21c) \quad & y_i \geq \gamma_{\min} \mu / C_0 \quad (i \in \mathcal{B}), & \lambda_i & \geq \gamma_{\min} \mu / C_0 \quad (i \in \mathcal{N}). \end{aligned}$$

Lemma 3.1 implies that the limit point $(z^*, \hat{\lambda}, y^*)$ defined in (9) has

$$(22) \quad \hat{\lambda}_i > 0, \quad i \in \mathcal{B}, \quad y_i^* = -g_i(z^*) > 0, \quad i \in \mathcal{N}.$$

The second result is a general one that relates different components of the solutions of the systems (14), (15), (16), (18), and (19).

LEMMA 3.2. (cf. [10, Lemma 5.2]) *Suppose that the standing assumptions are satisfied, and consider a general system of the form*

$$(23) \quad \begin{bmatrix} D_z f & Dg^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \bar{t} \\ \bar{u} \\ \bar{v} \end{bmatrix} = \begin{bmatrix} d^1 \\ d^2 \\ d^3 \end{bmatrix},$$

where (d^1, d^2, d^3) represents an arbitrary right-hand side. Then there are constants $\hat{\delta} \in (0, 1)$ and $C_1 > 0$ such that for all $(z, \lambda, y) \in \mathcal{S}(\hat{\delta})$, the solution of (23) satisfies

$$(24) \quad \|\bar{t}\| \leq C_1 (\|(d^1, d^2, d^3)\| + \mu \|\bar{u}_B\|).$$

We can choose C_1 such that the same estimate holds if we replace $Dg(z)$ by $Dg(z^*)$ in the coefficient matrix in (23).

Proof. Note first from [10, Equation (78)] that

$$(25) \quad \|(z, \lambda) - (z^*, \bar{\pi})\| \leq C_{1,1} \mu, \quad \text{for all } (z, \lambda, y) \in \mathcal{S}(\delta_2),$$

for some positive constants $C_{1,1}$ and $\delta_2 \in (0, 1)$. Note too that as in the proof of [10, Lemma 5.2], we have for any $\delta \in (0, 1]$ that $\mu = O(\delta)$ for all $(z, \lambda, y) \in \mathcal{S}(\delta)$.

By eliminating the \bar{v} component from (23), we obtain

$$\begin{bmatrix} D_z f & Dg^T \\ -Dg & \Lambda^{-1} Y \end{bmatrix} \begin{bmatrix} \bar{t} \\ \bar{u} \end{bmatrix} = \begin{bmatrix} d^1 \\ d^2 + \Lambda^{-1} d^3 \end{bmatrix}.$$

We reduce the system further by eliminating \bar{u}_N to obtain

$$(26) \quad \begin{aligned} & \begin{bmatrix} (D_z f) + (Dg_N)^T \Lambda_N Y_N^{-1} (Dg_N) & (Dg_B)^T \\ -(Dg_B) & \Lambda_B^{-1} Y_B \end{bmatrix} \begin{bmatrix} \bar{t} \\ \bar{u}_B \end{bmatrix} \\ & = \begin{bmatrix} d^1 - (Dg_N)^T \Lambda_N Y_N^{-1} [d^2 + \Lambda_N^{-1} d^3] \\ d_B^2 + \Lambda_B^{-1} d_B^3 \end{bmatrix} \triangleq \begin{bmatrix} \bar{d}^1 \\ \bar{d}_B^2 \end{bmatrix}. \end{aligned}$$

Because of the bounds (21), we have for any $(z, \lambda, y) \in \mathcal{S}(1)$ that

$$(27) \quad \|\bar{d}^1\| \leq \|d^1\| + \|Dg_N\| (\|\Lambda_N Y_N^{-1}\| \|d^2\| + \|Y_N^{-1}\| \|d_N^3\|) \leq C_{1,2} \|(d^1, d^2, d^3)\|,$$

and

$$(28) \quad \|\bar{d}_{\mathcal{B}}^2\| \leq \|d_{\mathcal{B}}^2\| + \|\Lambda_{\mathcal{B}}^{-1}\| \|d_{\mathcal{B}}^3\| \leq C_{1,2} \|(d^1, d^2, d^3)\|.$$

for some constant $C_{1,2}$. By noting that $\Lambda_{\mathcal{N}} Y_{\mathcal{N}}^{-1} = O(\mu)$ and $\Lambda_{\mathcal{B}}^{-1} Y_{\mathcal{B}} = O(\mu)$, and that

$$\begin{aligned} D_z f(z, \lambda) - D_z f(z^*, \hat{\lambda}) &= O(\|(z, \lambda) - (z^*, \hat{\lambda})\|) = O(\delta), \\ Dg_{\mathcal{B}} - Dg_{\mathcal{B}}^* &= O(\|z - z^*\|) = O(\mu), \end{aligned}$$

for all $(z, \lambda, y) \in \mathcal{S}(\delta)$ and all $\delta \in (0, 1]$, we have by rearranging (26) that

$$\begin{aligned} & \begin{bmatrix} D_z f(z^*, \hat{\lambda}) & (Dg_{\mathcal{B}}^*)^T \\ -(Dg_{\mathcal{B}}^*) & 0 \end{bmatrix} \begin{bmatrix} \bar{t} \\ \bar{u}_{\mathcal{B}} \end{bmatrix} \\ &= \begin{bmatrix} \bar{d}^1 + [D_z f(z^*, \hat{\lambda}) - D_z f(z, \lambda)]\bar{t} - (Dg_{\mathcal{N}})^T \Lambda_{\mathcal{N}} Y_{\mathcal{N}}^{-1} (Dg_{\mathcal{N}})\bar{t} + (Dg_{\mathcal{B}}^* - Dg_{\mathcal{B}})\bar{u}_{\mathcal{B}} \\ \bar{d}_{\mathcal{B}}^2 - (Dg_{\mathcal{B}}^* - Dg_{\mathcal{B}})\bar{t} - \Lambda_{\mathcal{B}}^{-1} Y_{\mathcal{B}} \bar{u}_{\mathcal{B}} \end{bmatrix} \\ &= \begin{bmatrix} \bar{d}^1 \\ \bar{d}_{\mathcal{B}}^2 \end{bmatrix} + O((\delta + \mu)\|\bar{t}\|) + O(\mu\|\bar{u}_{\mathcal{B}}\|), \end{aligned}$$

for all $(z, \lambda, y) \in \mathcal{S}(\delta)$, where $\delta \in (0, \delta_2)$ and δ_2 is defined as in (25). By partitioning \bar{t} into its components in $\ker Dg_{\mathcal{B}}^*$ and $\text{ran}(Dg_{\mathcal{B}}^*)^T$, we have from Assumption 6 that \bar{t} is bounded in norm by a multiple of the right-hand side in the system above. Hence, by applying (27) and (28), we can define a constant $C_{1,3}$ such that

$$(29) \quad \|\bar{t}\| \leq C_{1,3} [\|(d^1, d^2, d^3)\| + \mu\|\bar{u}_{\mathcal{B}}\| + (\mu + \delta)\|\bar{t}\|].$$

We can choose $\hat{\delta}$ small enough that the coefficient of $\|\bar{t}\|$ on the right-hand side of (29) is smaller than 0.5 for all $\delta \in (0, \hat{\delta}]$. By rearranging this expression, we obtain

$$\|\bar{t}\| \leq 2C_{1,3} [\|(d^1, d^2, d^3)\| + \mu\|\bar{u}_{\mathcal{B}}\|],$$

which verifies (24).

The proof of the last statement is similar. \square

We can apply the lemma above to the systems (16) and (18) to obtain the following result.

LEMMA 3.3. (cf. [10, Lemma 5.2]) *Suppose that the standing assumptions are satisfied. Then there are constants $\hat{\delta} \in (0, 1)$ and $C_2 > 0$ such that for all $(z, \lambda, y) \in \mathcal{S}(\hat{\delta})$, the solutions $(\widehat{\Delta}z, \widehat{\Delta}\lambda, \widehat{\Delta}y)$ of (16) and $(\tilde{t}, \tilde{u}, \tilde{v})$ of (18) satisfy the following relations:*

$$(30a) \quad \|\widehat{\Delta}z\| \leq C_2 \mu (1 + \|\widehat{\Delta}\lambda_{\mathcal{B}}\|),$$

$$(30b) \quad \|\tilde{t}\| \leq C_2 \mu (1 + \|\tilde{u}_{\mathcal{B}}\|).$$

Proof. To prove these result, we need to verify only that the right-hand sides of (16) and (18) are bounded by a multiple of μ . By Lipschitz continuity (Assumption 5), the definitions (8) and (10), and the fact that $f(z^*, \bar{\pi}) = 0$ where $(z^*, \bar{\pi})$ is defined in (12), we can decrease $\hat{\delta}$ of Lemma 3.2 if necessary so that there is a constant $C_{2,1} > 0$ for which

$$\begin{aligned} & \|\bar{\eta}_f\| \\ & \leq \|f(z, \lambda) + (D_z f)(z^* - z) + Dg(z)^T(\bar{\pi} - \lambda) - f(z^*, \bar{\pi})\| + \|f(z, \lambda)\| \\ & \quad + \|Dg(z) - Dg(z^*)\| \|\bar{\pi} - \lambda\| \\ & \leq L\|(z, \lambda) - (z^*, \bar{\pi})\|^2 + \beta_{\max} \mu + L\|z - z^*\| \|\bar{\pi} - \lambda\| \\ (31) \quad & \leq C_{2,1} \mu, \quad \text{for all } (z, \lambda, y) \in \mathcal{S}(\hat{\delta}), \end{aligned}$$

where L denotes the Lipschitz constant of Assumption 5. (The radius $\hat{\delta}$ is chosen small enough that $\mathcal{S}(\hat{\delta})$ lies inside the neighborhood of Assumption 5.) For the second right-hand side component, we have simply that

$$\begin{aligned} \|\bar{\eta}_g\| &\leq \|r_g\| + \|g(z^*) - g(z) - Dg(z^*)(z^* - z)\| \\ &\leq \beta_{\max}\mu + L\|z^* - z\|^2 \\ (32) \quad &\leq C_{2,2}\mu, \quad \text{for all } (z, \lambda, y) \in \mathcal{S}(\hat{\delta}), \end{aligned}$$

for some constant $C_{2,2}$. For the remaining right-hand-side component in (16), we have trivially that

$$\|\Lambda Y e\|_1 = m\mu.$$

The results now follow immediately from Lemma 3.2. \square

The next result and others following make use of the positive diagonal matrix D defined by

$$(33) \quad D = \Lambda^{-1/2} Y^{1/2}.$$

From Lemma 3.1, there is a constant C_3 such that

$$(34) \quad \|D\| \leq C_3\mu^{-1/2}, \quad \|D^{-1}\| \leq C_3\mu^{-1/2},$$

for all $(z, \lambda, y) \in \mathcal{S}(1)$.

LEMMA 3.4. (cf. [10, Lemma 5.3]) *Suppose that the standing assumptions are satisfied. Then there are constants $\hat{\delta} > 0$ and $C_4 > 0$ such that*

$$(35) \quad \|D\widehat{\Delta}\lambda\| \leq C_4\mu^{1/2}, \quad \|D^{-1}\widehat{\Delta}y\| \leq C_4\mu^{1/2},$$

for all $(z, \lambda, y) \in \mathcal{S}(\hat{\delta})$.

Proof. Let $\hat{\delta}$ be the smaller of those defined in Lemmas 3.2 and 3.3, but decreased if necessary to ensure that

$$(36) \quad (z, \lambda, y) \in \mathcal{S}(\hat{\delta}) \Rightarrow \mu \leq 1.$$

The proof closely follows that of [10, Lemma 5.3], but we spell out the details here because the analytical techniques are also needed in a later result (Theorem 3.9).

Recall the splitting (17) of the step $(\widehat{\Delta}z, \widehat{\Delta}\lambda, \widehat{\Delta}y)$ into components $(\tilde{t}, \tilde{u}, \tilde{v})$ and $(\tilde{t}', \tilde{u}', \tilde{v}')$ defined by (18) and (19), respectively. By multiplying the last block row in (19) by $\Lambda^{-1/2}Y^{-1/2}$ and using (33), we find that

$$(37) \quad D\tilde{u}' + D^{-1}\tilde{v}' = -(\Lambda Y)^{1/2}e.$$

Using (19) again, we obtain

$$(\tilde{u}')^T \tilde{v}' = -(\tilde{u}')^T (Dg^*)\tilde{t}' = (\tilde{t}')^T (D_z f)\tilde{t}' \geq 0,$$

since $D_z f$ is positive semidefinite by Assumption 1. Hence, by taking inner products of both sides in (37), we obtain

$$\|D\tilde{u}'\|^2 + \|D^{-1}\tilde{v}'\|^2 \leq \|(\Lambda Y)^{1/2}e\|^2 = m\mu,$$

and therefore

$$(38) \quad \|D\tilde{u}'\| \leq m^{1/2}\mu^{1/2}, \quad \|D^{-1}\tilde{v}'\| \leq m^{1/2}\mu^{1/2}.$$

For $(\tilde{t}, \tilde{u}, \tilde{v})$, the third block row in (18) implies that $D\tilde{u} = -D^{-1}\tilde{v}$. Therefore, we have

$$(39) \quad \begin{aligned} -(Dg^*)\tilde{t} - \tilde{v} = \bar{\eta}_g &\Rightarrow -\tilde{u}^T(Dg^*)\tilde{t} + \tilde{u}^T D^2 \tilde{u} = \tilde{u}^T \bar{\eta}_g \\ &\Rightarrow \|D\tilde{u}\|^2 = (\tilde{u})^T \bar{\eta}_g + (\tilde{t})^T \bar{\eta}_f - (\tilde{t})^T (D_z f) \tilde{t} \\ &\Rightarrow \|D\tilde{u}\|^2 \leq \|D\tilde{u}\| \|D^{-1}\bar{\eta}_g\| + \|\tilde{t}\| \|\bar{\eta}_f\|, \end{aligned}$$

where again we have used monotonicity of $D_z f$. Define the constant $C_{4,1}$ as

$$C_{4,1} = \max(C_2 C_{2,1}, C_2 C_{2,1} C_3, C_{2,2} C_3).$$

From (30b), (31), (34), and (36), we have for $(z, \lambda, y) \in \mathcal{S}(\hat{\delta})$ that

$$\|\tilde{t}\| \|\bar{\eta}_f\| \leq C_2 \mu (1 + \|\tilde{u}_B\|) C_{2,1} \mu \leq C_2 C_{2,1} \mu^2 (1 + C_3 \mu^{-1/2} \|D\tilde{u}\|) \leq C_{4,1} (\mu^2 + \mu^{1/2} \|D\tilde{u}\|).$$

From (32) and (34), we have

$$\|D^{-1}\bar{\eta}_g\| \leq C_3 \mu^{-1/2} C_{2,2} \mu \leq C_{4,1} \mu^{1/2}.$$

By substituting the last two bounds into (39), we obtain

$$\|D\tilde{u}\|^2 - 2C_{4,1} \|D\tilde{u}\| \mu^{1/2} - C_{4,1} \mu^2 \leq 0.$$

It follows from this inequality by a standard argument that

$$\|D\tilde{u}\| \leq C_{4,2} \mu^{1/2},$$

for some constant $C_{4,2}$ depending only on $C_{4,1}$, and $\hat{\delta}$. By combining this bound with (17) and (38), we obtain

$$\|D\widehat{\Delta\lambda}\| \leq \|D\tilde{u}\| + \|D\tilde{u}'\| \leq (C_{4,2} + m^{1/2}) \mu^{1/2},$$

and the first part of (35) follows if we define $C_4 = C_{4,2} + m^{1/2}$. Since $D^{-1}\tilde{v} = -D\tilde{u}$, the second part of (35) follows likewise. \square

Bounds on some of the components of $\widehat{\Delta\lambda}$ and $\widehat{\Delta y}$ follow easily from Lemma 3.4.

THEOREM 3.5. (cf. [10, Theorem 5.4]) *Suppose that the standing assumptions are satisfied. Then there are constants $\hat{\delta} \in (0, 1)$ and $C_5 > 0$ such that*

$$(40) \quad \|\widehat{\Delta\lambda}_{\mathcal{N}}\| \leq C_5 \mu, \quad \|\widehat{\Delta y}_B\| \leq C_5 \mu.$$

Proof. Let $\hat{\delta}$ be as defined in Lemma 3.4. From the definition (33) and the bounds (35), we have

$$\left| \left(\frac{y_i}{\lambda_i} \right)^{1/2} \widehat{\Delta\lambda}_i \right| \leq \|D\widehat{\Delta\lambda}\| \leq C_4 \mu^{1/2},$$

for any $i \in \mathcal{N}$. Hence, by using (21), we obtain

$$|\widehat{\Delta\lambda}_i| \leq \left(\frac{\lambda_i}{y_i}\right)^{1/2} C_4 \mu^{1/2} \leq \frac{C_0 \mu^{1/2}}{\gamma_{\min}^{1/2}} C_4 \mu^{1/2},$$

which proves that $\|\widehat{\Delta\lambda}_{\mathcal{N}}\| \leq C_5 \mu$ for an obvious choice of C_5 . The bound on $\widehat{\Delta y}_{\mathcal{B}}$ is derived similarly. \square

LEMMA 3.6. (cf. [10, Lemma 5.10]) *Let $\emptyset \neq \mathcal{J} \subset \mathcal{B}$ and $\emptyset \neq \mathcal{K} \subset \mathcal{N}$. If the two-sided projection of $D_z f(z, \lambda)$ onto $\ker Dg_{\mathcal{B}}^*$ is positive definite, then for $t \in \mathbb{R}^n$ and $\pi_{\mathcal{J}} \in \mathbb{R}^{|\mathcal{J}|}$, we have that*

$$(t, \pi_{\mathcal{J}}) \in \ker \begin{bmatrix} (D_z f) & (Dg_{\mathcal{J}}^*)^T \\ -Dg_{\mathcal{J}}^* & 0 \end{bmatrix}$$

if and only if $t = 0$ and $\pi_{\mathcal{J}} \in \ker(Dg_{\mathcal{J}}^*)^T$. In addition, we have that

$$\dim \ker \begin{bmatrix} (D_z f) & (Dg_{\mathcal{J}}^*)^T & 0 \\ -Dg_{\mathcal{J}}^* & 0 & -I_{\mathcal{K}} \end{bmatrix} = \dim \ker(Dg_{\mathcal{J}}^*)^T.$$

Proof. This result differs from [10, Lemma 5.10] only in that z^* replaces z as the argument of $Dg(\cdot)$. The proof is essentially unchanged. \square

By Assumptions 5 and 6, the two-sided projection of $D_z f(z, \lambda)$ onto the kernel of $(Dg_{\mathcal{B}}^*)$ is positive definite for all (z, λ, y) sufficiently close to the limit point $(z^*, \hat{\lambda}, y^*)$ defined in (9). It follows from Lemma 3.6 and (22) that the set

$$(41) \quad \left\{ \begin{bmatrix} D_z f(z, \lambda) & (Dg_{\mathcal{B}}^*)^T & 0 \\ -(Dg_{\mathcal{B}}^*) & 0 & -I_{\mathcal{N}} \end{bmatrix} : \|(z, \lambda) - (z^*, \hat{\lambda})\| \leq \bar{\delta} \right\}$$

has constant column rank for some $\bar{\delta} > 0$.

THEOREM 3.7. *Suppose that the standing assumptions hold. Then there is a positive constant $\tilde{\delta}$ such that for all $(z, \lambda, y) \in \mathcal{S}(\tilde{\delta})$, we have that $(\widehat{\Delta z}, \widehat{\Delta\lambda}_{\mathcal{B}}, \widehat{\Delta y}_{\mathcal{N}})$ is the solution of the following convex quadratic program:*

$$(42) \quad \begin{aligned} & \min_{(\bar{t}, \bar{u}_{\mathcal{B}}, \bar{v}_{\mathcal{N}})} \quad \frac{1}{2} \|D_{\mathcal{B}\mathcal{B}} \bar{u}_{\mathcal{B}}\|^2 + \frac{1}{2} \|(D_{\mathcal{N}\mathcal{N}})^{-1} \bar{v}_{\mathcal{N}}\|^2, \\ & \text{subject to} \end{aligned}$$

$$\begin{bmatrix} D_z f & (Dg_{\mathcal{B}}^*)^T & 0 \\ -(Dg_{\mathcal{B}}^*) & 0 & -(I_{\mathcal{N}}) \end{bmatrix} \begin{bmatrix} \bar{t} \\ \bar{u}_{\mathcal{B}} \\ \bar{v}_{\mathcal{N}} \end{bmatrix} = \begin{bmatrix} \bar{\eta}_f - (Dg_{\mathcal{N}}^*)^T \widehat{\Delta\lambda}_{\mathcal{N}} \\ \bar{\eta}_g + I_{\mathcal{B}} \widehat{\Delta y}_{\mathcal{B}} \end{bmatrix}.$$

Moreover, there is a constant C_6 such that

$$(43) \quad \|(\widehat{\Delta\lambda}_{\mathcal{B}}, \widehat{\Delta y}_{\mathcal{N}})\| \leq C_6 \|(\bar{\eta}_f, \bar{\eta}_g, \widehat{\Delta\lambda}_{\mathcal{N}}, \widehat{\Delta y}_{\mathcal{B}})\|.$$

Proof. The value $\tilde{\delta} = \min(\bar{\delta}, \hat{\delta})$, with $\bar{\delta}$ from (41) and $\hat{\delta}$ from Theorem 3.5, suffices to prove this result. The technique of proof is by now familiar (it follows the proof of [10, Theorem 5.12] closely), and we omit the details. \square

At this point, we have proved the first estimate in (20), as we summarize in the following theorem.

THEOREM 3.8. *Suppose that the standing assumptions hold. Then there are constants $\tilde{\delta} \in (0, 1)$ and $C_7 > 0$ such that for any $(z, \lambda, y) \in \mathcal{S}(\tilde{\delta})$ we have*

$$\|(\widehat{\Delta z}, \widehat{\Delta \lambda}, \widehat{\Delta y})\| \leq C_7 \mu.$$

Proof. Let $\tilde{\delta}$ be as defined in Theorem 3.7. From Theorem 3.5, (31), and (32), we have for $(z, \lambda, y) \in \mathcal{S}(\tilde{\delta})$ that

$$\|(\bar{\eta}_f, \bar{\eta}_g, \widehat{\Delta \lambda}_{\mathcal{N}}, \widehat{\Delta y}_{\mathcal{B}})\| = O(\mu).$$

Hence, from (43) we have also that

$$\|(\widehat{\Delta \lambda}_{\mathcal{B}}, \widehat{\Delta y}_{\mathcal{N}})\| = O(\mu),$$

and it follows from (30a) that $\|\widehat{\Delta z}\| = O(\mu)$. \square

Our last result is concerned with the second estimate in (20) involving the relationship between (t, u, v) and $(\widehat{\Delta z}, \widehat{\Delta \lambda}, \widehat{\Delta y})$.

THEOREM 3.9. *Suppose that the standing assumptions hold. Then there are constants $\delta \in (0, 1)$ and $C_8 > 0$ such that*

$$\|(\widehat{\Delta z} - t, \widehat{\Delta \lambda} - u, \widehat{\Delta y} - v)\| \leq C_8 \mu,$$

for all $(z, \lambda, y) \in \mathcal{S}(\delta)$.

Proof. By taking differences of (14) and (16), we obtain

$$(44) \quad \begin{bmatrix} D_z f & (Dg)^T & 0 \\ -(Dg) & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \widehat{\Delta z} - t \\ \widehat{\Delta \lambda} - u \\ \widehat{\Delta y} - v \end{bmatrix} = \begin{bmatrix} (Dg - Dg^*)(\bar{\pi} - \lambda) + (Dg - Dg^*)^T \widehat{\Delta \lambda} \\ (Dg - Dg^*)(z^* - z) + (Dg^* - Dg)\widehat{\Delta z} \\ 0 \end{bmatrix}.$$

We have from (25), Lipschitz continuity of $Dg(\cdot)$ (Assumption 5), and Theorem 3.8 that there is a radius $\delta_4 \in (0, \tilde{\delta})$ such that

$$(45) \quad \begin{bmatrix} (Dg - Dg^*)(\bar{\pi} - \lambda) + (Dg - Dg^*)^T \widehat{\Delta \lambda} \\ (Dg - Dg^*)(z^* - z) + (Dg^* - Dg)\widehat{\Delta z} \\ 0 \end{bmatrix} = O(\mu^2), \quad \text{all } (z, \lambda, y) \in \mathcal{S}(\delta_4).$$

The remainder of the proof follows that of [10, Lemma 5.7]. By applying Lemma 3.2 to the system (44), and using the estimate (45), we have that there are constants $\delta_5 \in (0, \delta_4)$ and $C_{8,1} > 0$ such that

$$(46) \quad \|\widehat{\Delta z} - t\| \leq C_{8,1} \left(\mu^2 + \mu \|u_{\mathcal{B}} - \widehat{\Delta \lambda}_{\mathcal{B}}\| \right), \quad \text{all } (z, \lambda, y) \in \mathcal{S}(\delta_5).$$

Next, we note that the technique used in the second half of the proof of Lemma 3.4 can be used to prove that there is $\delta \in (0, \delta_5)$ such that

$$(47) \quad \|D(\widehat{\Delta \lambda} - u)\| = \|D^{-1}(\widehat{\Delta y} - v)\| \leq C_{8,2} \mu^{3/2}, \quad \text{all } (z, \lambda, y) \in \mathcal{S}(\delta),$$

where D is the diagonal scaling matrix defined in (33). Modifications are needed only to account for the different right-hand side estimate (45) and the different estimate (46) of $\|\widehat{\Delta z} - t\|$; we omit the details. From (34) and (47), it follows immediately that

$$\|\widehat{\Delta \lambda} - u\| \leq C_{8,2} C_3 \mu, \quad \|\widehat{\Delta y} - v\| \leq C_{8,2} C_3 \mu.$$

The final estimate for $(\widehat{\Delta z} - t)$ is obtained by substituting these expressions into (46). \square

COROLLARY 3.10. *Suppose that the standing assumptions hold. Then there are constants $\delta > 0$ and C_9 such that the affine-scaling step defined by (6) satisfies*

$$\|(\Delta z, \Delta \lambda, \Delta y)\| \leq C_9 \mu, \quad \text{all } (z, \lambda, y) \in \mathcal{S}(\delta).$$

Proof. We have from Theorems 3.8 and 3.9 that $(t, u, v) = O(\mu)$ for δ defined as in Theorem 3.9. Moreover, it follows directly from [10, Section 5.3] that $(t', u', v') = O(\mu)$, possibly after some adjustment of δ . Hence, the result follows from (13). \square

4. Conclusions. The result proved here explains the numerical experience reported in Section 7 of Ralph and Wright [10], in which the convergence behavior of our test problems seemed to be the same regardless of whether the active constraint Jacobian satisfied the constant-rank condition. We speculated in [10] about possible relaxation of the constant-rank condition and have verified in this article that, in fact, this condition can be dispensed with altogether.

Our results are possibly the first proofs of superlinear convergence in nonlinear programming without multiplier nondegeneracy or uniqueness.

Acknowledgments. We thank the referees and associate editor for their comments and suggestions.

REFERENCES

- [1] J. F. BONNANS, *Local study of Newton type algorithms for constrained problems*, in Optimization—Fifth French-German Conference, S. Dolecki, ed., no. 1405 in Lecture Notes in Mathematics, Springer-Verlag, 1989, pp. 13–24.
- [2] F. FACCHINEI, A. FISCHER, AND C. KANZOW, *On the accurate identification of active constraints*, tech. rep., Universita' di Roma "La Sapienza", 1996.
- [3] A. FISCHER, *Modified Wilson method for nonlinear programs with nonunique multipliers*, Technical Report, Institute of Numerical Mathematics, Technical University of Dresden, Germany, February 1997.
- [4] W. W. HAGER, *Convergence of Wright's stabilized SQP algorithm*, Technical Report, Department of Mathematics, University of Florida, Florida, January 1997.
- [5] P. T. HARKER AND J.-S. PANG, *Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications*, *Mathematical Programming*, 48 (1990), pp. 161–220.
- [6] R. D. C. MONTEIRO AND S. J. WRIGHT, *Local convergence of interior-point algorithms for degenerate monotone LCP*, *Computational Optimization and Applications*, 3 (1994), pp. 131–155.
- [7] R. D. C. MONTEIRO AND F. ZHOU, *On superlinear convergence of infeasible-interior-point algorithms for linear constrained convex programs*, Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, Georgia, July 1995 (revised April 1996).
- [8] J.-S. PANG, *Convergence of splitting and Newton methods for complementarity problems: An application of some sensitivity results*, *Mathematical Programming*, 58 (1993), pp. 149–160.
- [9] L. QI AND Z. WEI, *Constant positive linear independence, KKT points and convergence of feasible SQP methods*, Technical Report, School of Mathematics, The University of New South Wales, Australia, 1997.

- [10] D. RALPH AND S. J. WRIGHT, *Superlinear convergence of an interior-point method for monotone variational inequalities*, *Complementarity and Variational Problems: State of the Art*, SIAM, 1997, 345–385.
- [11] S. J. WRIGHT, *A path-following interior-point algorithm for linear and quadratic optimization problems*, *Annals of Operations Research*, 62 (1996), pp. 103–130.
- [12] ———, *Primal-Dual Interior-Point Methods*, SIAM, 1996.
- [13] ———, *Superlinear convergence of a stabilized SQP method to a degenerate solution*, Preprint ANL/MCS-P643-0297, MCS Division, Argonne National Laboratory, Ill., April 1997. To appear in *Computational Optimization and Applications*.
- [14] ———, *Modifying SQP for degenerate problems*, Preprint ANL/MCS-P699-1097, MCS Division, Argonne National Laboratory, Ill., October 1997.
- [15] ———, *Effects of finite-precision arithmetic on interior-point methods for nonlinear programming*, Preprint ANL/MPC-P705-0198, MCS Division, Argonne National Laboratory, Ill., January 1998.

Appendix: The Algorithm. The major computational operation in the algorithm is the repeated solution of $n + 2m$ -dimensional linear systems of the form

$$(48) \quad \begin{bmatrix} D_z f & Dg^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta \lambda \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_f(z, \lambda) \\ r_g(z, y) \\ -\Lambda Y e + \tilde{\sigma} \mu_k e \end{bmatrix},$$

where the centering parameter $\tilde{\sigma}$ lies in the range $[0, \frac{1}{2}]$. These equations are simply the Newton equations for the nonlinear system of equality conditions from (2), except for the $\tilde{\sigma}$ term. The algorithm searches along the direction $(\Delta z, \Delta \lambda, \Delta y)$ obtained from (48).

At each iteration, the algorithm performs a *fast step* along a direction obtained by solving (6) (or, equivalently, (48) with $\tilde{\sigma} = 0$). We choose the neighborhood Ω_{k+1} to be strictly larger than Ω_k (by appropriate choice of γ_{k+1} and β_{k+1}), thereby allowing a nontrivial step α_k to be taken along this direction without leaving Ω_{k+1} . If the fast step achieves at least a certain fixed decrease in μ , it is accepted as the new iterate. Otherwise, we reset $\Omega_{k+1} \leftarrow \Omega_k$ and define a safe step by solving (48) with $\tilde{\sigma}$ chosen in the range $[\bar{\sigma}, 1)$ for some constant $\bar{\sigma} \in (0, \frac{1}{2})$. We perform a backtracking line search along this direction, stopping when we identify a value of α_k that achieves a “sufficient decrease” in μ without leaving the set Ω_{k+1} .

The algorithm is parametrized by the following quantities whose roles are explained more fully in [10].

$$\begin{aligned} \chi \in (0, 1), \quad \bar{\sigma} \in (0, \tfrac{1}{2}), \quad \bar{\alpha} \in (0, 1], \quad \kappa \in (0, 1), \quad \hat{\tau} \in (0, 1), \\ \beta_{\min} > 0, \quad \beta_{\max} = \beta_{\min} \exp(3/2), \quad 0 < \gamma_{\min} < \gamma_{\max} \leq \tfrac{1}{2}, \\ \bar{\gamma} \in (0, \tfrac{1}{2}), \quad \rho \in (0, \min((\tfrac{1}{2}\bar{\gamma})^{1/\hat{\tau}}, 1 - \kappa)), \end{aligned}$$

where $\exp(\cdot)$ is the exponential function. The constants β_{\min} and γ_{\max} are related to the starting point (z^0, λ^0, y^0) as follows:

$$\lambda_i^0 y_i^0 \geq \gamma_{\max} \mu_0, \quad \|r_f(z^0, \lambda^0)\| \leq \beta_{\min} \mu_0, \quad \|r_g(z^0, y^0)\| \leq \beta_{\min} \mu_0.$$

The main algorithm is as follows.

```

 $t_0 \leftarrow 0; \gamma_0 \leftarrow \gamma_{\max}; \beta_0 \leftarrow \beta_{\min};$ 
for  $k = 0, 1, 2, \dots,$ 
  if  $\mu_k = 0$ 
    terminate with solution  $(z^k, \lambda^k, y^k);$ 

```

```

 $(z^{k+1}, \lambda^{k+1}, y^{k+1}) \leftarrow \mathbf{fast}(z^k, \lambda^k, y^k, t_k, \gamma_k, \beta_k);$ 
if  $\mu_{k+1} \leq \rho\mu_k$ 
   $t_{k+1} \leftarrow t_k + 1;$ 
   $\gamma_{k+1} \leftarrow \gamma_{\min} + \bar{\gamma}^{t_{k+1}}(\gamma_{\max} - \gamma_{\min}); \beta_{k+1} \leftarrow (1 + \bar{\gamma}^{t_{k+1}})\beta_k;$ 
else
   $t_{k+1} \leftarrow t_k;$ 
   $(z^{k+1}, \lambda^{k+1}, y^{k+1}) \leftarrow \mathbf{safe}(z^k, \lambda^k, y^k, t_k, \gamma_k, \beta_k);$ 
   $\gamma_{k+1} \leftarrow \gamma_k; \beta_{k+1} \leftarrow \beta_k;$ 
end for.

```

Although we may calculate both a fast step and a safe step in the same iteration, the coefficient matrix in (48) is the same for both steps, so the coefficient matrix is factored only once.

The safe-step procedure is defined as follows.

```

safe( $z, \lambda, y, t, \gamma, \beta$ ):
  choose  $\tilde{\sigma} \in [\bar{\sigma}, \frac{1}{2}]$ ,  $\alpha^0 \in [\bar{\alpha}, 1]$ ;
  solve (48) to find  $(\Delta z, \Delta \lambda, \Delta y)$ ;
  choose  $\alpha$  to be the first element in the sequence  $\alpha^0, \chi\alpha^0, \chi^2\alpha^0, \dots$ ,
  such that the following conditions are satisfied:

```

$$\begin{aligned}
 \lambda_i(\alpha)y_i(\alpha) &\geq \gamma\mu(\alpha); \\
 \|r_f(z(\alpha), \lambda(\alpha))\| &\leq \beta\mu(\alpha); \\
 \|r_g(z(\alpha), y(\alpha))\| &\leq \beta\mu(\alpha); \\
 \mu(\alpha) &\leq [1 - \alpha\kappa(1 - \tilde{\sigma})]\mu
 \end{aligned}$$

```

return  $(z(\alpha), \lambda(\alpha), y(\alpha))$ .

```

The fast step routine is described next.

```

fast( $z, \lambda, y, t, \gamma, \beta$ ):
  solve (48) with  $\tilde{\sigma} = 0$  to find  $(\Delta z, \Delta \lambda, \Delta y)$ ;
  set  $\tilde{\gamma} = \gamma_{\min} + \bar{\gamma}^{t+1}(\gamma_{\max} - \gamma_{\min})$ ; set  $\tilde{\beta} = (1 + \bar{\gamma}^{t+1})\beta$ ;
  define

```

$$\alpha^0 = 1 - \frac{\mu^{\tilde{\gamma}}}{\bar{\gamma}^t};$$

```

if  $\alpha^0 \leq 0$  return( $z, \lambda, y$ );
  choose  $\alpha$  to be the first element in the sequence  $\alpha^0, \chi\alpha^0, \chi^2\alpha^0, \dots$ ,
  such that the following conditions are satisfied:

```

$$\begin{aligned}
 \lambda_i(\alpha)y_i(\alpha) &\geq \tilde{\gamma}\mu(\alpha); \\
 \|r_f(z(\alpha), \lambda(\alpha))\| &\leq \tilde{\beta}\mu(\alpha); \\
 \|r_g(z(\alpha), y(\alpha))\| &\leq \tilde{\beta}\mu(\alpha);
 \end{aligned}$$

```

return  $(z(\alpha), \lambda(\alpha), y(\alpha))$ .

```